Climate Services 9 (2018) 44–56

Contents lists available at ScienceDirect

Climate Services

journal homepage: www.elsevier.com/locate/cliser

Dynamical and statistical downscaling of seasonal temperature forecasts in Europe: Added value for user applications

R. Manzanas [a,*], J.M. Gutiérrez [a], J. Fernández [b], E. van Meijgaard [c], S. Calmanti [d], M.E. Magariño [b], A.S. Cofiño [b], S. Herrera [b]

[a] Meteorology Group, Instituto de Física de Cantabria (CSIC-UC), Santander 39005, Spain
[b] Meteorology Group, Dpto. de Matemática Aplicada y Ciencias de la Computación, Universidad de Cantabria, Santander 39005, Spain
[c] Royal Netherlands Meteorological Institute (KNMI), PO Box 201, 3730 AE De Bilt, The Netherlands
[d] Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA, SSPT-MET-CLIM), Via Anguillarese 301, 00123 Roma, Italy

ARTICLE INFO

Article history:
Available online 16 June 2017

Keywords:
Dynamical downscaling
Statistical downscaling
Seasonal forecasting
Europe
Heatwaves

ABSTRACT

This work describes the results of a comprehensive intercomparison experiment of dynamical and statistical downscaling methods performed in the framework of the SPECS (http://www.specs-fp7.eu) and EUPORIAS (http://www.euporias.eu) projects for seasonal forecasting over Europe, a region which exhibits low-to-moderate seasonal forecast skill. We considered a 15-member hindcast provided by the EC-EARTH global model (similar to ECMWF System 4, but using bias corrected SST) for the period 1991–2012. In particular, we focus on summer mean temperature and evaluate the added value of downscaling for representation of the local climatology (mean values and extremes), improvement of model skill and performance in particular heatwave episodes. Whereas the suitability of dynamical downscaling for reducing the orographic biases of the global model depends on the region and model considered, statistical downscaling can systematically reduce errors in different order moments, from the mean to the extremes (as represented by the 95th percentile here). However, both dynamical and statistical methods lead to similar skill patterns with about the same overall performance as the global model, which shows higher values in south-eastern Europe. Therefore, no relevant added value is found in terms of model skill improvement. Finally, when focusing on the heatwaves of 2003, 2006, 2010 and 2012, the limitations of the global model to detect these hot episodes are inherited by all dynamical and statistical downscaling methods so no added value is neither found in this aspect. This work provides, to our knowledge, the largest and most comprehensive intercomparison of statistical and dynamical downscaling for seasonal forecasting over Europe.

Practical Implications

Dynamical and statistical downscaling methods allow transferring the coarse biased seasonal predictions from global ocean–atmosphere coupled models to the regional/local spatial scales required in impact studies, providing thus actionable products which properly represent the local features of interest. However, whereas both approaches have been extensively used and critically assessed in climate change studies, their added value for seasonal forecasting is not well understood yet, and comprehensive intercomparison studies over Europe are still lacking.

In this work we focus on this problem and consider several representative dynamical and statistical methods—which have been used in the framework of the SPECS (http://www.specs-fp7.eu) and EUPORIAS (http://www.euporias.eu) projects—to downscale the seasonal forecasts of summer temperature over Europe from a state-of-the-art global model. We evaluate the (possible) added value of downscaling, both dynamical and statistical, in terms of representativeness of the local climatology (mean values and extremes), improvement of model skill and performance in particular extreme episodes (2003, 2006, 2012 and 2012 heatwaves). This comprehensive intercomparison provides therefore key information for European stakeholders focused on different socio-economic sectors.

* Corresponding author.
E-mail address: rmanzanas@ifca.unican.es (R. Manzanas).

http://dx.doi.org/10.1016/j.cliser.2017.06.004
2405-8807/

Our results show that, whereas the suitability of dynamical downscaling for reducing the orographic biases of the global model depends on the region and model considered, statistical downscaling can systematically reduce errors in different order moments, from the mean to the extremes (as represented by the 95th percentile here), providing thus more realistic climate information than global models do. This can have important practical implications for different user applications in a range of sectors such as agriculture, energy, health or tourism, for which the use of realistic seasonal forecasts is increasingly growing. For the case of dynamical downscaling, it is worth to mention that sophisticated quantile mapping techniques have been recently applied to regional models so that they match the local observation's statistics grid box by grid box. This approach (dynamical downscaling followed by bias adjustment) has become a common practice in different user applications—particularly in the context of multi-decadal climate change, where regional climate projections are readily available—and it is preferable to the direct bias adjustment of the global model outputs, particularly when local (point-wise) information is required.

Differently, no relevant added value is found in terms of model skill improvement, neither for dynamical nor for statistical methods. Both downscaling approaches lead to similar skill patterns (evaluated by means of the ROC Skill Score here) with about the same overall performance as the global model, which shows low-to-moderate skill over most of the continent—the highest skill being located over south-eastern Europe and for cold events. Similar conclusions have been also reported in Nikulin et al. (2018) for East Africa, a region with different skill and climate characteristics. Nevertheless, note that the ROC Skill Score (ROCSS) is not sensitive to mean errors and thus, other bias-dependent performance measures such as the Root Mean Square Error (RMSE) or the Continuous Ranked Probability Skill Score (CRPSS) could still indicate that some added value may be obtained from downscaling. As for the ROCSS, no added value is found here in terms of reliability, neither for dynamical nor for statistical methods, all of them yielding similar results, overall comparable to the ones provided by the global model. This suggests that other strategies rather than downscaling, such as the use of probability calibration techniques (Primo et al., 2009), might be needed to improve the reliability of the global seasonal forecasts over Europe. However, since they work based on interannual probabilities, these techniques require long hindcasts (over 30 years) for proper calibration and validation. This should be taken into account when defining the hindcasts feeding climate services such as Copernicus (http://www.copernicus.eu).

Finally, when focusing on particular heatwaves (2003, 2006, 2010 and 2012), dynamical and statistical methods are shown to inherit the limitations of the global model, which fails in detecting these anomalously hot episodes, both in magnitude (much lower than observed) and location. Despite this, recent works have found that models can reach some skill for prediction of these events if soil moisture is properly initialized. Yet, further research is still needed in order to assess the performance of raw and downscaled seasonal climate data to forecast extreme indicators such as hot/cold spells, which may be relevant for different practical applications.

In summary, beyond the reduction of global model biases, our results indicate that there is no clear signal of added value for downscaling, neither dynamical nor statistical, for seasonal forecasts of summer temperature over Europe. Moreover, in agreement with a previous intercomparison study performed in south-eastern USA (Schoof et al., 2009), there is no clear indication on which of the two approaches is preferable. In this regard, it is important to note the elevated requirements of dynamical downscaling (as compared to statistical one) in terms of computing resources and time. For instance, whereas the WRF model took on 190,000 core hours in the Altamira HPC facility—which is part of the Spanish Supercomputer Network (see acknowledgements),—the statistical methods considered required less than 6 core hours to downscale the whole experiment in a regular workstation. Therefore, dynamical downscaling experiments should be carefully designed in order to maximize the information gained from such expensive simulations. With respect to the statistical methods, we have considered in this work two daily Perfect Prognosis (PP) and one monthly Model Output Statistics (MOS) techniques. It is worth to highlight that, when applied under a leave-one-out cross-validation framework, the MOS technique was found to provide worse results than the PP ones for all validation aspects considered. However, if no cross-validation is performed, high artificial skill appears as a result of model overfitting, outperforming all other methods. This warns on the misuse of MOS methods for monthly/seasonal forecasting.

The results from this work constitute the most comprehensive to date intercomparison of dynamical and statistical downscaling for seasonal forecasts on a continental scale. However, it must be noticed that the conclusions drawn here are only for summer temperature over Europe, and may be not extensible to other variables, regions and seasons. Further investigation is still needed to provide a more conclusive overview on the merits and limitations of dynamical and statistical downscaling for seasonal forecasting.

Both the global and downscaled (dynamical and statistical) seasonal forecasts used in this work have been published for the ECOMS' community through the User Data Gateway (UDG: http://meteo.unican.es/udg-tds, *ecoms* catalog), which combines a THREDDS data server with web-services to manage datasets, catalogs, users and authorization protocols, leading to a practical tool to explore and access the available datasets (see Cofiño et al., 2018, for details).

1. Introduction

Global ocean–atmosphere coupled models are the primary tool used nowadays to generate seasonal climate forecasts. However, these models suffer from substantial biases (when compared with surface observations) and are unable to provide useful information at the regional or local spatial scales required in a number of sectors such as agriculture, energy, health, tourism or insurance (see Doblas-Reyes et al., 2013, and references therein). Therefore, an increasing interest has been focused in downscaling methods, which can transfer these global predictions to the regional/local spatial scales, providing thus actionable products representing the local features of interest. There are two main approaches to downscale climate information: statistical and dynamical. The latter is computationally very expensive, particularly in the case of seasonal forecasting where several members and initializations have to be downscaled. Moreover, the few existing experiences with dynamical models in seasonal forecasting have been mostly focused on tropical regions with moderate-to-high skill (see, e.g. Chou et al., 2005; Diro et al., 2012; Yoon et al., 2012; Diro, 2016). Thus, downscaling of seasonal forecasts has mainly relied on statistical methods (see Gutiérrez et al., 2013a, for a review), and studies over Europe, the target region in this work, are very scarce (Díez et al., 2005; Frías et al., 2010; Díez et al., 2011; Patarčić and Branković, 2012).

In contrast to the tropics, seasonal predictability over Europe is still limited (see, e.g., Doblas-Reyes et al., 2013, and references therein). However, the recent advances made with regard to the predictability of the NAO (see, e.g., Scaife et al., 2014) and the new potential sources for seasonal skill such as global warming, stratosphere-troposphere interactions and soil processes (see, e.g., Koster et al., 2010; Cohen and Jones, 2011) pave the way for better seasonal forecasts. As a result, there is a growing interest for this type of predictions, which are being demanded by different impact applications across Europe. In this work we focus on this promising region and present the results of an intercomparison experiment with dynamical and statistical downscaling methods carried out in the framework of the SPECS (http://www.specs-fp7.eu) and EUPORIAS (http://www.euporias.eu) projects. Therefore, this paper provides timely and useful information on the added value of the different downscaling approaches for seasonal forecasts, which is nowadays limited for European stakeholders.

Dynamical downscaling is based on regional models, which run on a relatively fine grid (e.g. 10–20 km) over a limited domain (e.g. Europe) initialized and driven at the boundaries by the coarse global model outputs (see, e.g., Laprise, 2008). These models are able to generate regional predictions for a suite of climate variables, but still may suffer from significant biases which require post-processing with bias adjustment techniques before they can be used in impact applications (Yoon et al., 2012). In this work we consider three regional models (RACMO2, WRF and RegCM) which have been previously used over Europe in a series of multi-decadal climate change coordinated experiments such as EURO-CORDEX (Jacob et al., 2014).

Statistical downscaling is based on empirical relationships derived between a local observed predictand of interest (summer temperature in this case) and one or several suitable model predictors, either from reanalysis (in the Perfect Prognosis, PP, approach) or from global seasonal forecasting systems (Model Output Statistics, MOS, approach). In this work we consider two representative PP techniques (analogs and multiple linear regression), which have been previously calibrated over Europe in the framework of the COST action VALUE (Maraun et al., 2015), and one MOS regression technique. Note that statistical downscaling methods could in principle take advantage of atmospheric teleconnections by extending the predictor region well beyond the target region or by using different teleconnection indices as predictors. However, most of the teleconnections patterns responsible for seasonal variability (e.g. ENSO) have in general weak effect over Europe (see, e.g., Rust et al., 2015; Brands, 2017). Therefore, these ad hoc methods are left out in this work and deserve a separate study, focusing on variables and seasons for which teleconnections are stronger.

The relative merits and limitations of dynamical and statistical downscaling have been widely discussed in the literature (see, e.g,. Wilby and Wigley, 1997; Fowler et al., 2007; Maraun et al., 2010; Winkler et al., 2011) and it is nowadays recognized that both are complementary in many practical applications. Nevertheless, whereas both approaches have been extensively used and critically assessed in climate change studies (see, e.g., Murphy, 2000; Spak et al., 2007, for an example over Europe and North America, respectively), their added value for seasonal forecasting is not well understood yet, and there is a lack of comprehensive intercomparison studies over Europe—the few existing experiences are quite inconclusive or provide limited information for this region (see, e.g. Shukla and Lettenmaier, 2013).

Therefore, the main goal of this paper is to analyze the added value of dynamical and statistical downscaling over Europe, focusing on summer (JJA) mean temperature, which was one of the case studies selected by the Cross-Cutting Theme 3 (CCT3) in the SPECS project. To this aim, we consider a state-of-the-art global seasonal forecasting system, developed in the framework of the EUPORIAS project and based on the EC-EARTH model (more details in Nikulin et al., 2018) and downscale its 15-member hindcast for the period 1991–2012 over Europe using the different downscaling alternatives previously mentioned. In particular, we and evaluate the added value of downscaling considering three aspects: 1) representativeness of the local climatology (mean values and extremes), 2) improvement of model skill, and 3) performance in particular extreme episodes (2003, 2006, 2012 and 2012 heatwaves). Several verification metrics and indicators are considered to this aim. The results from this work constitute the most comprehensive to date intercomparison of dynamical and statistical downscaling for seasonal forecasts on a continental scale (note that a dynamical versus statistical comparison for downscaling of seasonal forecasts of precipitation have been done in Nikulin et al. (2018) for East Africa).

The paper is organized as follows. In Section 2 we introduce the data and experimental framework used, describing the statistical and dynamical methods considered. The results obtained are presented and discussed through Section 3. Finally, the most important conclusions are given in Section 4.

2. Methods and data

2.1. Observations

In this work we used daily gridded observations from E-OBS—in particular its version 11 (Haylock et al., 2008),—at a regular 0.25° resolution, as the reference dataset to train the different downscaling methods and validate the raw and downscaled seasonal predictions. Although observational uncertainty may be an important source of uncertainty in downscaling and validation studies (Nikulin et al., 2018), in this work we do not consider this factor due to the high quality and density of temperature records over Europe from which E-OBS has been built.

2.2. Global seasonal predictions

The global seasonal predictions considered in this work have been produced in the framework of the EUPORIAS project, using EC-EARTH version 3.1 (Hazeleger et al., 2010). This model was run in atmosphere-only mode with an effective horizontal resolution of about 75 km, forced by the bias-corrected SST pattern from the corresponding ECMWF System 4 forecasts (more details in Nikulin et al., 2018). Here we downscale the full hindcast available, consisting of 15 members initialized on the first of May and providing data from May to September for the period 1991 to 2012. All members were initialized from identical atmosphere and soil conditions (including soil moisture), which were directly taken from the ECMWF System 4, and they only differ in the SST/sea ice initial state and forcing. As indicated in the introduction, this work provides dynamically and statistically downscaled summer (JJA) mean temperatures for the whole Europe during the 22-year period 1991–2012.

2.3. Dynamical downscaling

In this work, we considered three regional models used in SPECS to downscale the EC-EARTH global model. On the one hand, RACMO2 (run by KNMI) and WRF (run by CSIC-UCAN), which were run over the EURO-CORDEX domain at 0.22° resolution. On the other hand, RegCM (run by ENEA), which was run over the Med-CORDEX domain (Ruti et al., 2016) at 30 km resolution. They are described next.

RACMO2 (van Meijgaard et al., 2012) is a hydrostatic model primarily built on the semi-lagrangian dynamical kernel of HIRLAM 6.3.7 (Undén et al., 2003) and the physical parameterization package of the ECMWF IFS, basically cy31r2 (similar to the cycle used in ERA-Interim; ECWMF, 2007, but containing a few modifications and extensions). The version used here is the same as used in the EURO-CORDEX experiments, employing 40 hybrid coordinate full vertical levels and using the following physical parameterizations: a short wave radiation transfer scheme with 6 spectral intervals (Fouquart and Bonnel, 1980); RRTM 16 intervals for long wave radiative transfer (Mlawer et al., 1997); the Eddy-Diffusivity Mass

Flux scheme (Siebesma et al., 2007), extended with prognostic turbulent kinetic energy (Lenderink and Holtslag, 2004), to describe dry and moist turbulent mixing processes and shallow convection in the boundary layer; the HTESSEL scheme (Balsamo et al., 2009) with revised land surface hydrology (introduced in cycle33r1; ECWMF, 2009) to describe land surface/soil processes; a bulk mass flux scheme originally developed by Tiedtke (1989) with many extensions to describe deep cumulus convection (Nordeng, 1994; Neggers et al., 2009); and a prognostic cloud scheme to describe stratiform clouds and large-scale precipitation (Tiedtke, 1993; Tompkins et al., 2007). Vegetation maps are inferred from ECOCLIMAP (Champeaux et al., 2005). Initial soil moisture was obtained by remapping the EC-EARTH soil wetness index onto the own regional model grid, so that soil moisture values fall in the range governed by local soil moisture wilting point and field capacity.

The Weather Research and Forecasting (WRF; Skamarock et al., 2008) modelling system features a non-hydrostatic dynamical core, which solves the fully-compressible Euler equations in flux form. The WRF model used in this work is ARW-WRF-v3.4.1, applying the configuration used by UCAN in EURO-CORDEX (code WRF341I). This configuration considers 30 full eta vertical levels and the following physical parameterizations: CAM scheme (Collins et al., 2004) for long and short wave radiation; the Yonsei University (YSU; Hong et al., 2006) non-local closure planetary boundary layer (PBL) scheme; and the Noah land surface model (Chen and Dudhia, 2001). For moist processes, the WRF single-moment scheme with 5 microphysics species (WSM5; Hong et al., 2006) was used, along with the Kain-Fritsch cumulus scheme to account for unresolved convection (Kain, 2004).

The RegCM modeling system (Giorgi et al., 2012) is a hydrostatic, compressible, sigma-p vertical coordinate model run on an Arakawa B-grid in which wind and thermodynamical variables are horizontally staggered. The version used for this work is the RegCM-4.3.5.6, which considers 18 sigma-p levels and the following physical parameterizations: the CCM3 radiative transfer scheme (Kiehl et al., 1996); a modified version of the Holtslag parameterization for the planetary boundary layer (Giorgi et al., 2012); and the Noah land surface model. For moist processes, the RegCM configuration considered here uses the cumulus convection model described in Grell et al. (1994), with a Fritsch-Chappell scheme for unresolved convection. The resolved scale precipitation is modeled with the SUBEX parameterization (Pal et al., 2000), whose parameters have been optimized for the specific model domain after the configuration adopted by Artale et al. (2010). In contrast to RACMO2 and WRF, note that the domain covered by RegCM does not include the northernmost regions of the continent.

2.4. Statistical downscaling

As previously mentioned, statistical downscaling relies on models/algorithms which link the coarse-resolution global simulated predictors with the local observed predictand over the area of interest (see, e.g., von Storch et al., 1993). These models/algorithms are first trained/calibrated using historical data of both coarse model predictors and local predictands for a sufficiently long period (usually a few decades) and then applied to new (e.g., future) model predictors to obtain the corresponding local downscaled predictands. According to the nature of predictors used to train the different models, two approaches for statistical downscaling exist, namely Perfect Prognosis (PP) and Model Output Statistics (MOS) (see Marzban et al., 2006, for an interesting discussion on this). Whereas quasi-observed predictors from reanalysis are used in PP (see, e.g., Burger and Chen, 2005; Haylock et al., 2006; Fowler et al., 2007; Hertig and Jacobeit, 2008; Sauter and Venema, 2011; Gutiérrez et al., 2013b), predictors are directly taken from the global model in MOS (see, e.g., Mo and Straus; Sokol, 2003; Kang et al., 2004; Marzban et al., 2006; Vannitsem and Nicolis, 2008). Nevertheless, and despite this conceptual difference, both PP and MOS the models are calibrated at an event-wise (time-series) level; that is, making use of the temporal (e.g. day-to-day/season-to-season) correspondence between predictors and predictands (E-OBS temperature in this case). Thus, since the day-to-day correspondence with the observations is preserved in the case of reanalysis, PP techniques can be applied on a daily, monthly or seasonal basis, whereas MOS techniques require working at monthly or seasonal time-scales. Here we used two PP and one MOS technique, which are later described.

To calibrate the two PP techniques, daily predictors from ERA-Interim reanalysis (Dee et al., 2011) at a 2° horizontal resolution were used. Afterwards, the resulting statistical models were fed with the corresponding predictors from the EC-EARTH model (interpolated to the same 2° grid) to obtain the downscaled seasonal predictions. In order to properly harmonize the reanalysis and the global seasonal forecast data, a simple local scaling correction was applied to the latter. In particular, for every large-scale model predictor, monthly mean values were adjusted towards the corresponding climatological reanalysis values (computed over the entire period of study 1991–2012), gridbox by gridbox. This way, we avoid problems that may arise due to the global model mean biases.

One of the assumptions of the PP approach is that the predictors used should be well represented by both reanalysis and the global model (Wilby et al., 2004). Moreover, Manzanas et al. (2017) showed recently that statistical downscaling methods can yield significant skill improvements in cases for which the predictor variables considered are better predicted by the global model than the target predictand and have a strong link with local weather. Thus, to assess the suitability of some key predictor variables typically used for statistical downscaling of temperatures, Fig. 1 shows their interannual correlation between EC-EARTH and ERA-Interim. With the exception of temperature at 500 hPa (T500), the highest correlations are found for two-meter temperature (T2) and mean

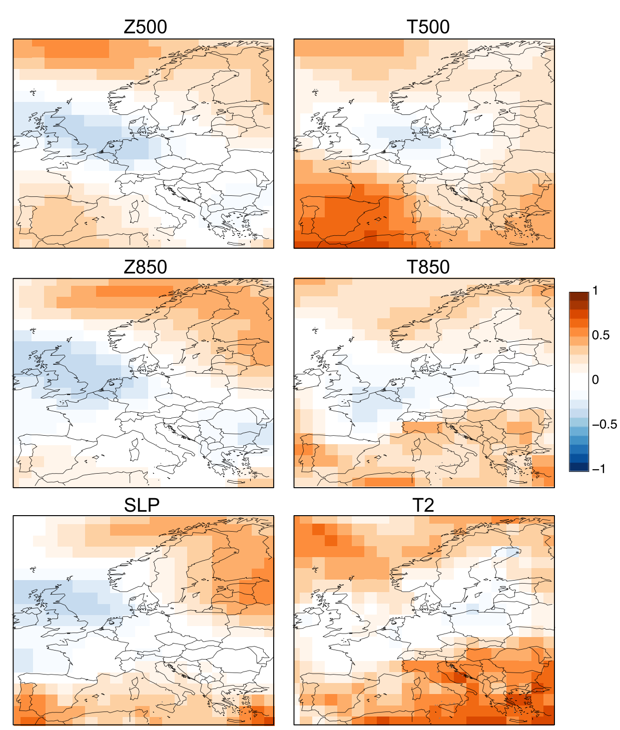

Fig. 1. Interannual correlation between EC-EARTH and ERA-Interim for a number of key predictor variables typically used for statistical downscaling of temperatures.

sea level pressure (SLP), especially over southern Europe. Based on these results, as well as on the work done in Gutiérrez et al. (2013b), we considered a combination of T2 and SLP as predictors in this work. These two variables account for thermodynamic and circulation effects and explain a major part of the variability of the predictand (note that T500 is not really explicative for surface temperature). Moreover, they allow to exploit the skill of the global model to the greatest extent.

Building on the comprehensive work done in the COST action VALUE (Maraun et al., 2015), the two PP methods used here were independently applied over the eight PRUDENCE zones (see, e.g., García-Díez et al., 2015), which were slightly modified here to cover the entire Europe. Moreover, in order to filter-out the high-frequency variability which may be not properly linked to the local-scale, instead of the raw T2 and SLP fields, we used the 15 leading principal components (PCs). PCs (Preisendorfer, 1988) were obtained, both for the reanalysis and for the seasonal forecasts, by projecting the corresponding standardized fields onto the Empirical Orthogonal Functions obtained from the reanalysis, which were computed simultaneously on the two predictor variables, considering the joined vector of standardized fields. The number of PCs retained, which explain about the 70–90% of the predictor variance across the different zones, was selected as a trade-off between model parsimony and goodness-of-fit.

The first of the two PP techniques considered (referred to as PP-ANA hereafter) is based on the popular analogue technique (Lorenz, 1969), which estimates the local downscaled values corresponding to a particular atmospheric configuration (as given by the forecasting model) from the local observations corresponding to a set of similar (or analog) atmospheric configurations within a historical catalog formed by reanalysis. Here, a deterministic version of the technique (Zorita et al., 1995; Cubasch et al., 1996) which considers the closest analog—in terms of the euclidean distance—was used. In spite of its simplicity, the analog technique performs as well as more sophisticated ones (Zorita and von Storch, 1999) and has been applied in several previous studies to downscale temperatures in the context of seasonal forecasting (see, e.g., Frías et al., 2010).

The second PP technique is a multiple linear regression (PP-MLR henceforward), an extension of simple linear regression which attempts to model the relationship between two or more explanatory variables (e.g. predictors) and a response variable (e.g. predictand) by fitting a linear equation to observed data. The fit is determined by minimizing the sum of the residuals between the regression line and the observed data. A detailed description on the theory of this technique is provided by Helsel and Hirsch (2002). Regression-based methods have also been used in previous works to downscale seasonal forecasts of temperature (see, e.g., Pavan et al., 2005).

In addition to the two PP techniques above described, which work on a daily basis, we also considered a MOS implementation of the multiple linear regression (MOS-MLR hereafter) working at a monthly basis. For consistency, this technique relied on the same predictors considered for the PP ones (T2 and SLP). In particular, for each E-OBS gridbox, we considered as predictors the monthly aggregated values at the four nearest EC-EARTH gridboxes. Note that, as opposite to PP methods, MOS techniques are trained with small data sample (e.g. 22 and 66 points for seasonal and monthly implementations, respectively, for the 22-year hindcast available for this work). Therefore, cross-validation is critical in this case to avoid artificial skill due to overfitting (see, e.g. Robertson et al., 2012). To assess the effect of cross-validation on the results obtained, the above presented MOS-MLR technique was applied with and without cross-validation. In the former case, a leave-one-out (Lachenbruch and Mickey, 1968) scheme was considered, in which each year was separately considered for test, keeping all the remaining ones for calibration. Note that this was not done for the PP methods since they are trained and tested with different datasets (reanalysis and seasonal forecasts) and, thus, cross-validation is not necessary (see a detailed discussion on this in Manzanas et al., 2017).

3. Results and discussion

In this section we evaluate the added value of downscaling in terms of representativeness of the local climatology (mean values and extremes) and improvement of model skill. Note that global and dynamically downscaled forecasts have been interpolated to the 0.25° E-OBS grid, using nearest neighbors, before validation.

3.1. Representation of mean and extreme values

First, we analyze the biases of the different downscaling methods as compared to the biases found in the coarse predictions of the EC-EARTH global model (second row of Fig. 2), using the E-OBS observations as reference (first row). As a result of its limited spatial resolution and the corresponding miss-representation of important local features (as compared to E-OBS), EC-EARTH exhibits strong biases in the mean and extreme (95th percentile: P95) values, particularly over regions of complex topography, such as the Alps and the Carpathians. In this regard, it is worth to note that simple statistical corrections such as the standard lapse rate (see, e.g., Sheridan et al., 2010) method may help to reduce these orography-related biases. Likewise, sophisticated bias adjustment techniques have been recently applied with the same purpose, calibrating the model outputs gridbox by gridbox to match local observation statistics (Yoon et al., 2012). However, these techniques can deteriorate the temporal structure (and skill) of the model (Manzanas et al., 2017), so the dynamical and statistical downscaling methods analyzed in this work should, in general, be preferable alternatives for reducing the global model biases.

Fig. 2 (rows 3 to 7) shows the biases for the three dynamical models considered in this work. The results from RACMO2 (for both mean and extreme values) exhibit small biases. In agreement with Kotlarski et al. (2014), the biases for the mean are mostly cold, especially in the western Alps. With respect to the P95, biases are both cold and hot, depending on the region. The strong warm bias over the northern coast of Africa (a common feature for all models) is most likely due to the scarce and unreliable observations which entered E-OBS in this area. As compared to the EC-EARTH, note that a clear added value is obtained for this model, which appropriately represents the local statistical properties of observations. The WRF model exhibits strong spatially-systematic cold biases over vast parts of the continent, mainly over central Europe, which is in agreement with the results obtained with 'perfect' (reanalysis) boundary conditions in previous studies (see, e.g., García-Díez et al., 2013, 2015; Katragkou et al., 2015). However, the spatial pattern is more uniform (not so orography-driven) than the one corresponding to the global EC-EARTH model. As such, very simple post-processing correction approaches could be easily applied. For instance, here we rescaled the results of the WRF model by considering its mean spatial bias, yielding significantly better results (fifth row), although with significant biases over yet.

RegCM presents a strong warm bias over eastern Europe, both in the mean (up to more than 3°C in many areas) and in the P95. Whereas the former is partially alleviated by means of the same spatial rescaling applied to WRF, the latter remains mostly the same. This positive bias of RegCM differs from the negative one obtained in previous studies (see, e.g., Artale et al., 2010; Patarčić and Branković, 2012). This may be explained by the particular configuration adopted here for the BATS land-surface scheme

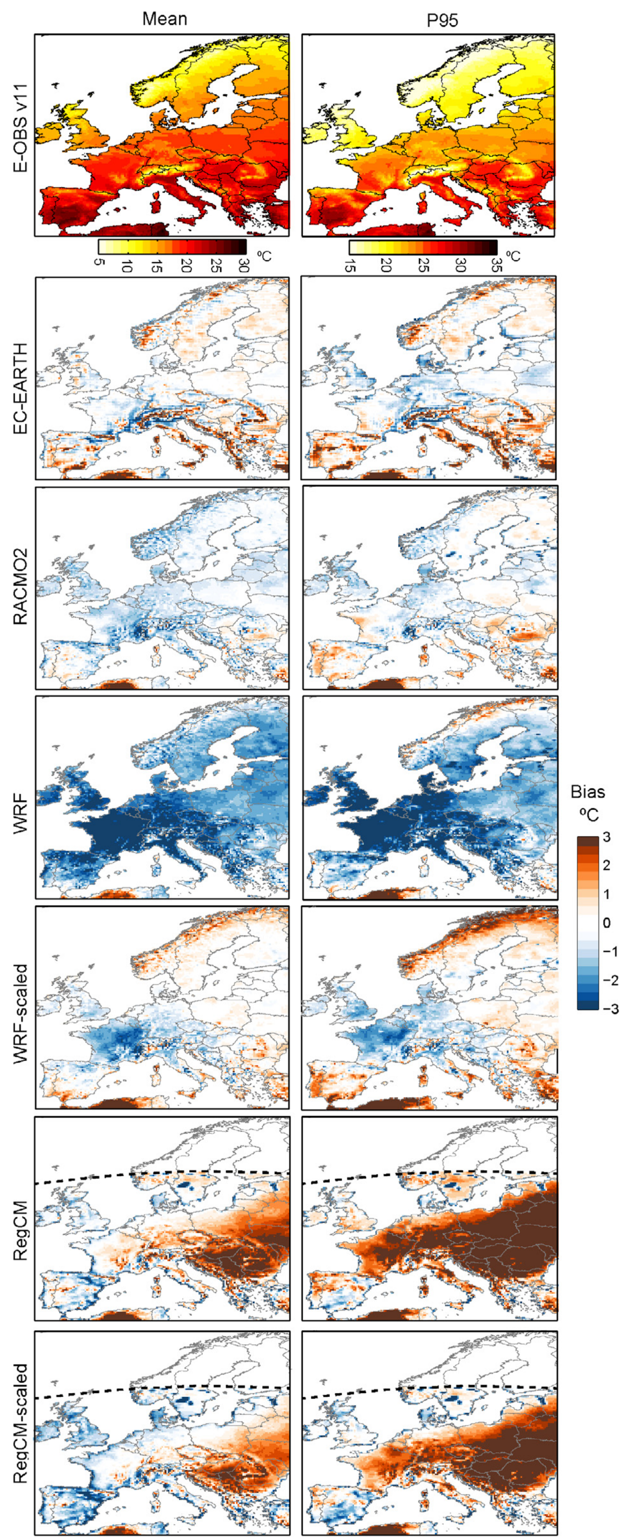

Fig. 2. Mean climatology and 95th percentile (in columns) of JJA mean temperature for 1991–2012. First row shows the observed values, as represented by E-OBS. Rows 2–7 show the biases of the EC-EARTH and the different dynamical downscaling products.

(Dickinson et al., 1993). or instance, with respect to the configuration used in Artale et al. (2010), the values of the minimum stomatal resistance for the main land-use types over eastern Europe (mixed crop/farming and deciduous broad-leaf tree) have been increased from 45 s m^{-1} and 60 s m^{-1} to 180 s m^{-1} and 200 s m^{-1}, respectively. This was done to decrease the release of latent heat (which is especially relevant in summer) with the idea of reducing the cold bias found by these authors. However, the results obtained here have finally led to an over-compensation of this error, which indicates that further adjustments are still needed to obtain optimal performances for this model.

Differently to the global model, for which biases are found to be mostly related to a misrepresented orography, these results indicate that biases in the regional models seem to be more process-related. Consequently, the spatial patterns can largely vary across the regional models (see, e.g., the opposite biases found for WRF and RegCM over France and eastern Europe, respectively). Therefore, and despite it helps to considerably reduce the orographic biases of the global model (both in the mean and in the P95), the suitability of dynamical downscaling depends in general on the region and model considered. For some cases, significant biases are still present in the downscaled results and the application of bias adjustment techniques may be a necessary post-processing step previous to using these results for impact studies. This approach (dynamical downscaling followed by bias adjustment) has become a common practice in different user applications.

Nevertheless, as expected by construction, all the statistical downscaling methods exhibit in general very small biases (Fig. 3), especially for the mean (note that either PP or MOS methods are calibrated towards the observed climate). Although still small, biases are slightly larger for extreme temperatures (P95), which are not directly taken into account in the calibration process (notice that the bias for the P95 can not be computed for the case of the MOS-MLR since this technique provides only monthly data). Indeed, for some regions, biases for P95 are comparable to those exhibited by the RACMO2 model (the best among the regional models). These results indicate the ability of statistical downscaling to systematically reduce errors in different order moments, from the mean to the P95, providing thus more realistic climate information than global models do.

All the previous results proof that downscaling (and especially statistical downscaling) provides a clear added value for user applications in those cases where regional/local calibrated information

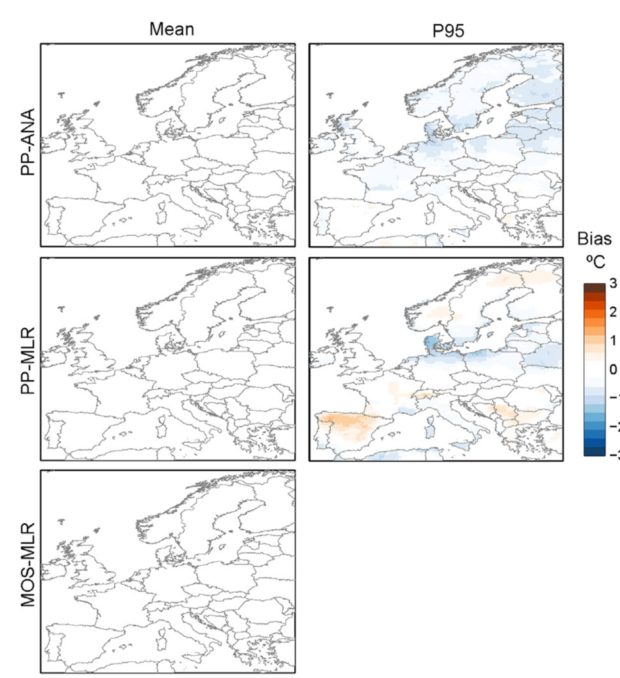

Fig. 3. Biases of the mean and 95th percentile (in columns) of JJA mean temperature for 1991–2012 for the different statistical downscaling methods.

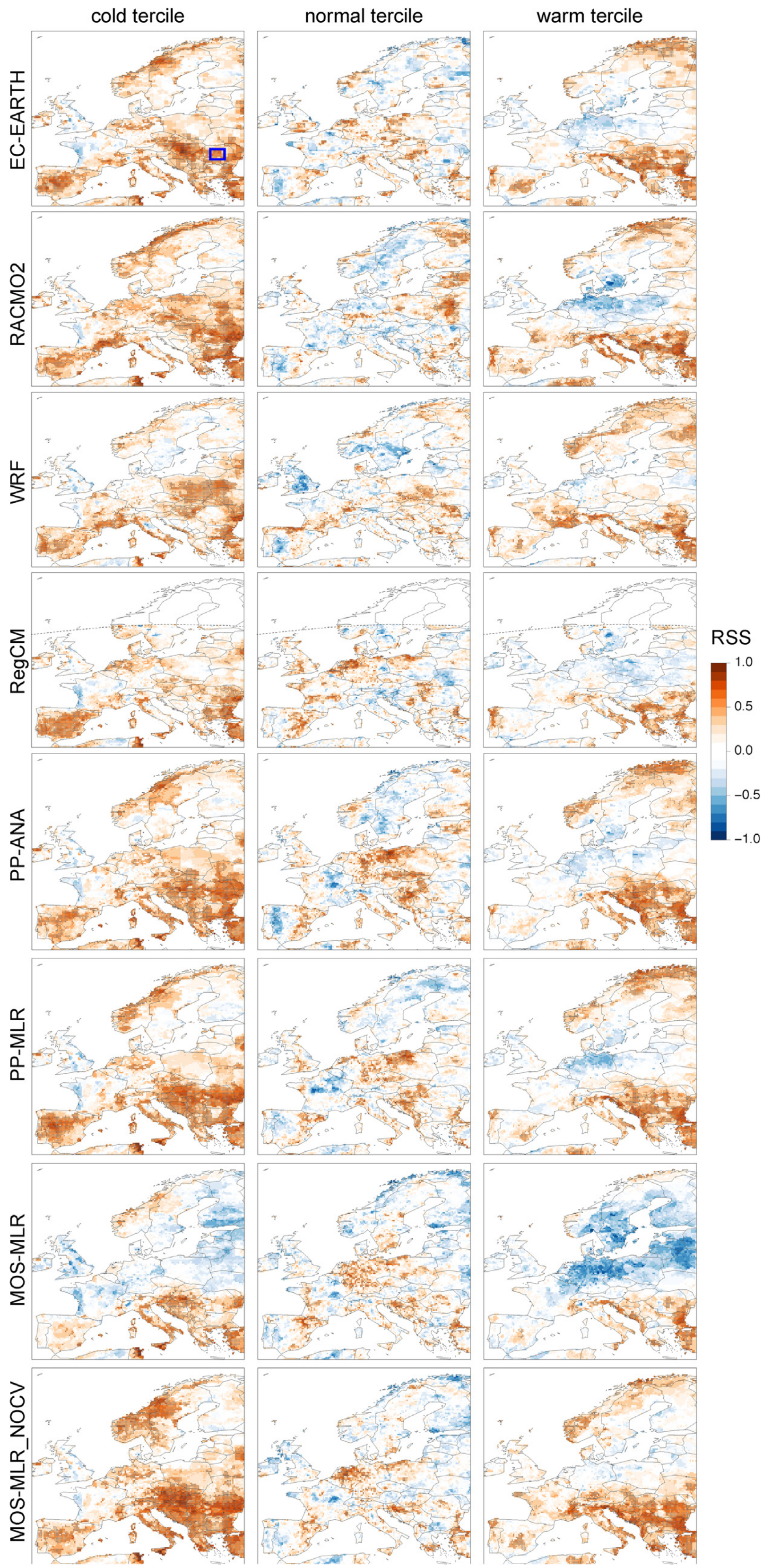

Fig. 4. ROCSS found for EC-EARTH, the dynamical models and the statistical downscaling methods considered (in rows) for the cold, normal and warm tercile (left, central and right column, respectively). Significant ($\alpha = 0.1$) values are indicated with a black dot. The blue rectangle represents the area considered in Fig. 5. For illustrative purposes, the results for MOS-MLR without cross-validation are also shown in the last row.

is required to obtain sector specific climate indices and/or run impact models. Note that, even for those impact applications which require point-wise seasonal forecasts matching the local observations' statistics, downscaling is preferable to bias adjustment techniques, whose suitability may be questioned due to the scale gap between global model predictions and observations (Maraun, 2016).

3.2. Skill and reliability

In this section we assess the performance of the downscaling methods focusing on different aspects of forecast quality: accuracy and reliability. The global direct forecasts from EC-EARTH are considered as a benchmark to evaluate the potential added value of downscaling. Note that, even though all the dynamical and statistical experiments performed here (with the exception of the MOS-MLR technique) provided daily downscaled forecasts, the validation metrics described next were applied on the seasonal averages (i.e. interannual series) computed from these daily values. In all cases, E-OBS was used as reference observations.

On the one hand, accuracy measures the degree to what forecasts and observations agree. Here we use the ROC Skill Score (ROCSS) as a measure of accuracy for probabilistic forecasts of each of the three terciles (cold, normal and warm). The ROCSS is computed as $2A - 1$, where A is the area under the ROC curve. The ROC curve is obtained for binary probabilistic forecasts—different probability thresholds as considered—as the proportion of occurrences that were correctly forecast versus the proportion of non-occurrences that were incorrectly forecast. Therefore, the ROCSS ranges between 1 ($A = 1$: perfect forecast system) and -1 ($A = 0$: perfectly bad forecast system), with a zero value ($A = 0.5$) indicating no skill compared with a climatological prediction. The ROCSS is a reasonable first choice to communicate the value of a forecast to the end-users (see, e.g. Manzanas et al., 2014) and it is recommended by the Lead Centre for the Standardized Verification System of Long Range Forecasts (http://www.bom.gov.au/wmo/lrfvs/index.html).

Fig. 4 shows the ROCSS obtained for EC-EARTH, the dynamical models and the statistical methods for the cold, normal and warm terciles. Black dots indicate statistically significant ($\alpha = 0.1$) values. Although some regional differences are found across the different methods, in general, neither dynamical nor statistical downscaling greatly modify the skill pattern exhibited by EC-EARTH, which exhibits the best results for south-eastern Europe (note the correspondence with Fig. 1, which shows that the highest skill for the predictors used is located in this region), and in particular for the cold tercile. Specially interesting is to highlight the important role that cross-validation plays for the case of the MOS-MLR. Whereas this technique is found to provide the highest skill if no cross-validation is considered, its results turn clearly worse than those obtained for the two PP methods when the leave-one-out framework is applied. This evidences the necessity of proper cross-validation in monthly/seasonal MOS techniques and warns on the credibility of some of the good results found in previous work for this type of techniques.

To further analyze the effect of the downscaling approaches applied, Fig. 5 shows, for the illustrative region indicated with a blue rectangle in the upper-left panel of Fig. 4 (for which there is some general positive skill), the tercile plots (Manzanas et al., 2014) for the EC-EARTH and the different downscaled products. Each plot displays, year by year, the predicted probabilities (as obtained from the frequencies of the 15-member ensemble) for each of the three terciles (in rows) in a white-to-orange colored scale, together with the observed tercile (black circles). This figure allows to visually inspect to what extent downscaling can modify the interannual variability of the probabilities given by the global model. Whereas there is a general agreement between the global model and the different downscaling methods (particularly high for certain cases, such as for the cold tercile in 1996), there are some exceptions for which downscaling can introduce changes in the global prediction. For instance, most of the methods significantly increase the probability of the warm tercile for 2003 towards the observed value. Alternatively, there are also cases for which downscaling can wrongly modify the prediction of the global model (e.g. the normal tercile in 2000). Also, this figure shows that part of the skill attained at this region may be attributable to the existing simultaneous trends in both the observations and the predictions (higher probabilities and tercile occurrences most frequent for cold/warm terciles at the beginning/end of the period). As a matter of fact, the ROCSS patterns shown in Fig. 4 turn weaker when removing these trends (not shown)—in the case of the predictions, the trend of the ensemble mean is used to detrend each individual member.

On the other hand, reliability measures how closely the forecast probabilities of a certain event (e.g. occurrence of a particular tercile) correspond to the actual chance of observing that event. To assess this important aspect in seasonal forecasting, Weisheimer and Palmer (2014) proposed a user-oriented scale with five reliability categories: *perfect* (green), *still very useful* (blue), *marginally useful* (yellow), *not useful* (orange) and *dangerously useless* (red). These categories are based on the corresponding reliability diagrams; in particular on the relative position of the reliability line and the uncertainty range around it, which is estimated by bootstrapping. This classification has been recently modified by Manzanas et al. (2017), who, within the *marginally useful* category, differentiated those cases in which the reliability line lies within the skill region, assigning to this new category (*marginally useful +*) the dark yellow color (see the original work for further details on the classification).

Reliability categories were computed for each of the eight extended PRUDENCE zones (the same which were considered for the calibration and application of the PP techniques). Within each region, observations and probabilistic predictions for all the grid boxes were joined into a single time-series in order to increase

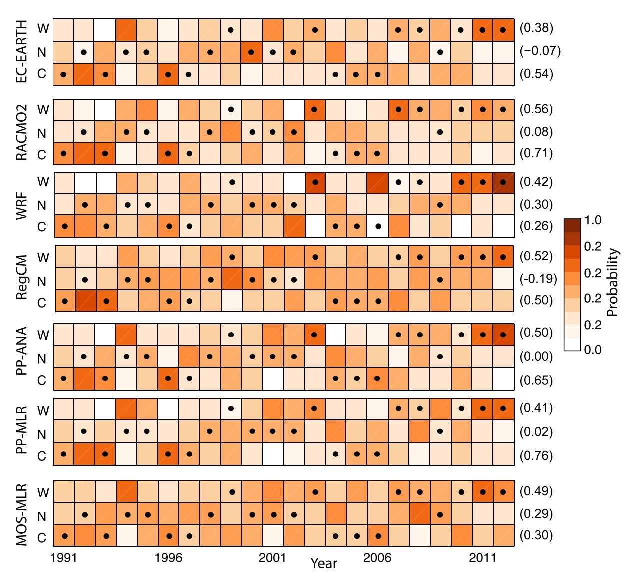

Fig. 5. Tercile plots (see the text) for EC-EARTH and the downscaled forecasts, spatially averaged over the blue area highlighted in Fig. 4. Each plot displays, year by year, the predicted probabilities (color scale) for each of the three terciles (in rows, W: warm, N: normal, C: cold), together with the observed tercile (black circles). Numbers within parenthesis on the right indicate the ROCSS for each tercile.

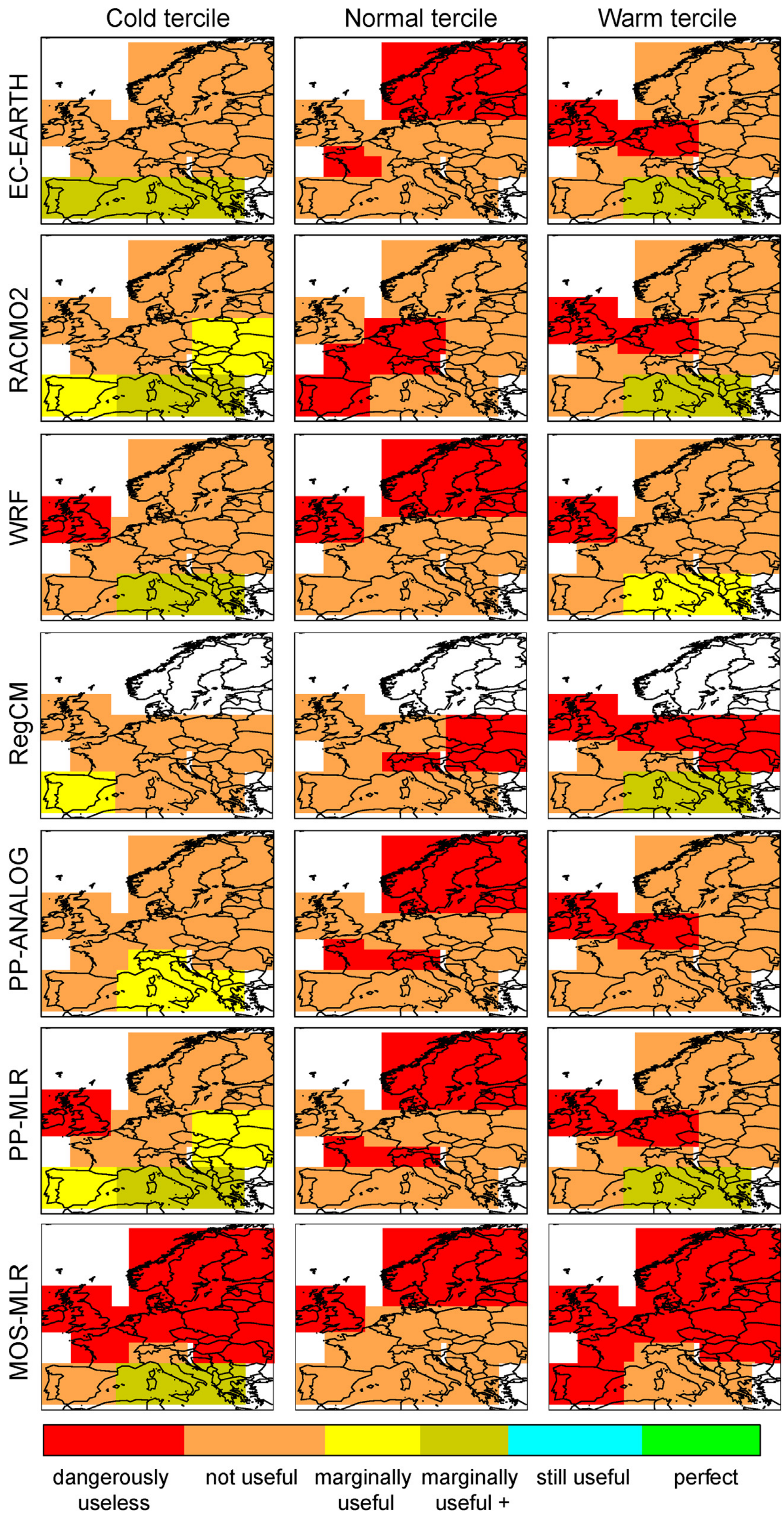

Fig. 6. Reliability categories obtained for the EC-EARTH, the dynamical models and the statistical methods considered (in rows) for the cold, normal and warm tercile (left, central and right column, respectively), over the eight PRUDENCE zones. Note that the northernmost regions are not shown for RegCM since they are not totally covered in this model's domain. See Manzanas et al. (2017) for details on the methodology used.

the sample size. Fig. 6 shows the results obtained for the EC-EARTH and the downscaled products for each tercile. With the exception of the Mediterranean region, EC-EARTH exhibits *not useful* or *dangerously useful* categories for most of the regions. Moreover, as for the ROCSS, neither dynamical nor statistical downscaling greatly alter the results found for the global model. As revealed from the

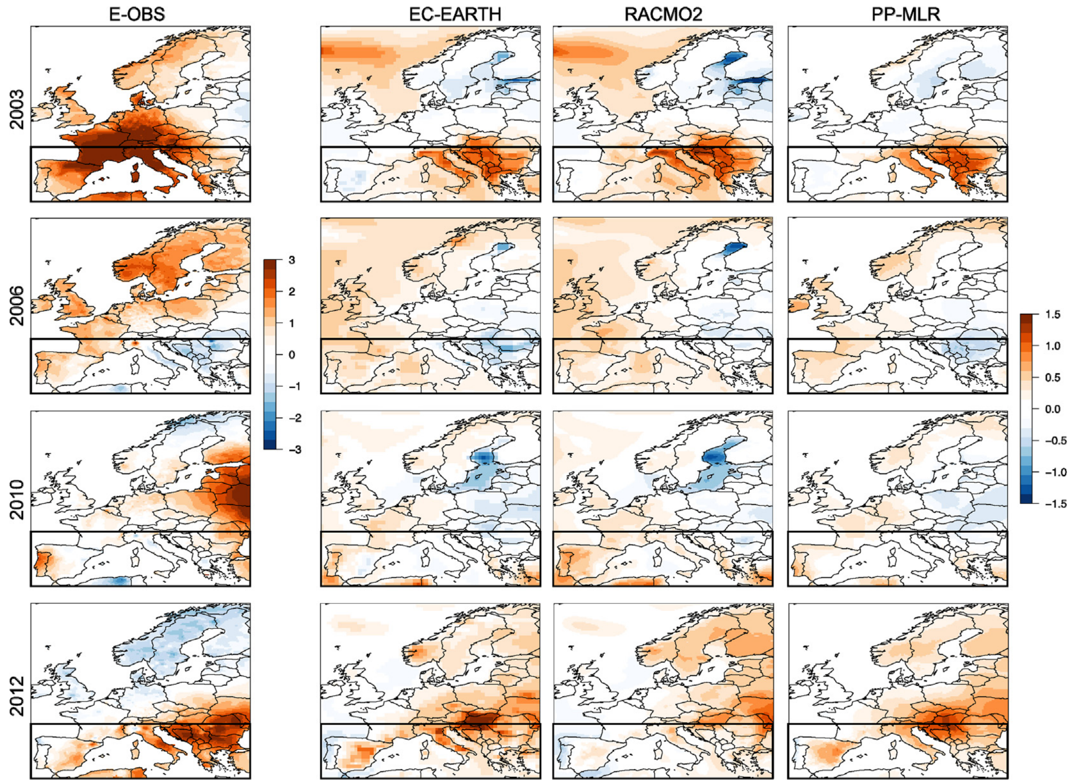

Fig. 7. JJA mean temperature anomalies for 2003, 2006, 2010 and 2012 (in rows), as represented by E-OBS, EC-EARTH, RACMO2 and PP-MLR (in columns).

corresponding reliability diagrams (not shown), the differences found across the distinct downscaling methods and regions are due to spurious changes of category, which are related to the bootstrapping process followed to estimate the uncertainty around slope of the reliability line. Differently from Weisheimer and Palmer (2014) and Manzanas et al. (2017), who considered a confidence interval of 75% and 50%, respectively, here we used a more conservative 90%. If this confidence interval is relaxed, the results obtained tend to improve over northern Europe for the cold tercile, passing from *not useful* to *marginallly useful* and/or *marginally useful* + categories in many cases. Yet, these results suggest that other approaches rather than downscaling, such as probability calibration techniques (Primo et al., 2009), might be needed to improve the reliability of the global seasonal forecasts over Europe. Note that the comments already made on the poorer performance of the MOS technique (as compared to the PP ones) also apply here.

3.3. Heatwave events

The analysis performed in the previous section indicates the overall performance of the global model and the different downscaling alternatives considered over the period 1991–2012. However, it is also important to assess their usefulness for predicting particular conditions which can have important impacts on different sectors. For instance, since the early 2000's, a number of heatwaves have affected various parts of Europe, with significant socioeconomic impacts such as heat-related mortality and financial losses due to crop failure or wildfire damages. For this reason, we focus in this section in the four severe heatwaves experienced across many parts of Europe during the summers of 2003, 2006, 2010 and 2012, which were selected by the SPECS Cross-Cutting Theme 3 (CCT3), in alignment with the "seasonal warm/dry case study" experiment. Thus, the results presented here could be intercompared and evaluated from a wider (process-based) perspective within the SPECS CCT3 framework. For instance, in accordance with previous works (see, e.g., Hirschi et al., 2011; Patarčić and Branković, 2012), Eden et al. (2014) and Ardilouze et al. (2016) have highlighted the importance of proper soil moisture conditions for detection of the above heatwaves.

Fig. 7 shows the JJA mean temperature anomalies for these four years (in rows), as represented by E-OBS, EC-EARTH, RACMO2 and PP-MLR (in columns). The two latter were selected as representative of the dynamical and the statistical downscaling approach, respectively. The large regional anomalies observed, which reached 4°C in central Europe (2003), northern Europe (2006), Russia (2010) and south-eastern Europe (2012), were not properly captured by EC-EARTH. In fact, not only the magnitude of the anomaly (which is much lower in the global model), but also the spatial pattern was wrongly predicted in most of the cases. Moreover, both dynamical and statistical downscaling approaches yield very similar results to EC-EARTH, indicating that scarce added value should be expected from downscaling for detection of heatwaves. Despite this, if we focus only on the Mediterranean region indicated by the black rectangle, there is still some room for optimism, with EC-EARTH (and the corresponding downscaled forecasts) exhibiting a reasonable degree of agreement with E-OBS. Yet, further research is still needed in order to assess the performance of raw and downscaled seasonal climate data to forecast extreme indicators such as hot/cold spells, which may be relevant for different practical applications.

4. Conclusions

This work assesses the (possible) added value of downscaling for seasonal forecasts of summer temperature over Europe. To do this, we consider several dynamical regional models and statistical methods which have been carried out in the framework of the SPECS (http://www.specs-fp7.eu) and EUPORIAS (http://www.euporias.eu) projects to downscale a state-of-the-art global model. The following three aspects are evaluated: 1) representativeness of the local climatology (mean values and extremes), 2) improvement of model skill, and 3) performance in particular extreme episodes (2003, 2006, 2012 and 2012 heatwaves).

First, our results show that, whereas the suitability of dynamical downscaling for reducing the orographic biases of the global model depends on the region and model considered, statistical downscaling can systematically reduce errors in different order moments, from the mean to the extremes (as represented by the 95th percentile here), providing thus more realistic climate information than global models do. This can have important practical implications for different user applications in a range of sectors such as agriculture, energy, health or tourism, for which the use of realistic seasonal forecasts is increasingly growing. However, it is worth to notice at this point that, as a result of the calibration step based on observations, statistical downscaling is expected to reduce biases by construction. Thus, for a fairer comparison, other factors such as bias adjustment of the final results or the use of better soil moisture initialization strategies might be considered for the case of dynamical models.

Second, no relevant added value is found in terms of model skill improvement, neither for dynamical nor for statistical methods. Both downscaling approaches lead to similar skill patterns (evaluated by means of the ROC Skill Score here) with about the same overall performance as the global model, which shows low-to-moderate skill over most of the continent (the highest skill being located over south-eastern Europe and for cold events). However, the ROC Skill Score (ROCSS) is not sensitive to mean errors and thus, other bias-dependent performance measures such as the Root Mean Square Error (RMSE) or the Continuous Ranked Probability Skill Score (CRPSS) could still indicate that some added value may be obtained from downscaling. As for the ROCSS, no added value is found in terms of reliability neither for dynamical nor for statistical methods, all of them yielding similar results, overall comparable to the ones provided by the global model.

Third, when focusing on particular heatwaves (2003, 2006, 2010 and 2012), dynamical and statistical methods are shown to inherit the limitations of the global model, which fails in detecting these anomalously hot episodes.

In summary, beyond the reduction of global model biases (which is particularly evident for the case of statistical downscaling), we have not found clear signal of the added value of downscaling, neither dynamical nor statistical, for seasonal forecasts of summer temperature over Europe. Moreover, there is no clear indication on which of the two approaches is preferable. In this regard, as compared to statistical downscaling, it is important to note the elevated requirements of dynamical downscaling in terms of computing resources and time. With respect to the statistical methods, we have considered for this work two daily Perfect Prognosis (PP) and one monthly Model Output Statistics (MOS) techniques. The MOS technique was found to provide worse results than the PP ones for all validation aspects here considered when applied under a leave-one-out cross-validation framework. However, as a result of the artificial skill that appears, this technique becomes the best one if no proper cross-validation is considered. This warns on the misuse of MOS methods for monthly/seasonal forecasting.

The results from this work constitute the most comprehensive to date intercomparison of dynamical and statistical downscaling for seasonal forecasts on a continental scale. However, it must be noticed that the conclusions drawn here are only for summer temperature over Europe, and may be not extensible to other variables, regions and seasons. Further investigation is still needed to provide a more conclusive overview on the merits and limitations of dynamical and statistical downscaling of seasonal forecasts.

Acknowledgments

This study was supported by the SPECS and EUPORIAS projects, funded by the European Commission through the Seventh Framework Programme for Research under grant agreements 308378 and 308291, respectively. We are also grateful to the E-OBS dataset from the EU-FP6 project ENSEMBLES and the data providers in the ECA&D project. One of the authors (EvM) wants to thank Michael Kolax from SMHI (Norrköping, Sweden) for making available the full EC-EARTH hindcast ensemble for dynamical downscaling at KNMI. Finally, for the WRF simulations, the authors acknowledge the access provided to the Altamira Supercomputer at the Institute of Physics of Cantabria (IFCA-CSIC), member of the Spanish Supercomputing Network (http://grid.ifca.es/wiki/Supercomputing).

References

Ardilouze, C., van Meijgaard, E., Prodhomme, C., Batté, L., Déqué, M., vanden Hurk, B., 2016. Predictability of summer warm events: lessons learnt from three case studies over Europe. Technical Report. SPECS Technical Notes 5. URL: <http://www.specs-fp7.eu/sites/default/files/u1/SPECSTechnicalNoteCaseStudies.pdf>.

Artale, V., Calmanti, S., Carillo, A., Dell'Aquila, A., Herrmann, M., Pisacane, G., Ruti, P. M., Sannino, G., Struglia, M.V., Giorgi, F., Bi, X., Pal, J.S., Rauscher, S., 2010. An atmosphere–ocean regional climate model for the mediterranean area: assessment of a present climate simulation. Clim. Dyn. 35, 721–740. http://dx.doi.org/10.1007/s00382-009-0691-8.

Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., vanden Hurk, B., Hirschi, M., Betts, A. K., 2009. A revised hydrology for the ECMWF Model: verification from field site to terrestrial water storage and impact in the integrated forecast system. J. Hydrometeorol. 10, 623–643. http://dx.doi.org/10.1175/2008JHM1068.1.

Brands, S., 2017. Which ENSO teleconnections are robust to internal atmospheric variability? Geophys. Res. Lett. 44, 1483–1493. http://dx.doi.org/10.1002/2016GL071529.

Burger, G., Chen, Y., 2005. Regression-based downscaling of spatial variability for hydrologic applications. J. Hydrol. 311, 299–317. http://dx.doi.org/10.1016/j.jhydrol.2005.01.025. URL: <http://www.sciencedirect.com/science/article/pii/S0022169405000740>.

Champeaux, J.L., Masson, V., Chauvin, F., 2005. ECOCLIMAP: a global database of land surface parameters at 1 km resolution. Meteorol. Appl. 12, 29–32. http://dx.doi.org/10.1017/S1350482705001519.

Chen, F., Dudhia, J., 2001. Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: Model implementation and sensitivity. Mon. Weather Rev. 129, 569–585. http://dx.doi.org/10.1175/1520-0493(2001) 1290569:CAALSH2.0.CO;2.

Chou, S.C., Bustamante, J.F., Gomes, J.L., 2005. Evaluation of Eta Model seasonal precipitation forecasts over South America. Nonlinear Processes Geophys. 12, 537–555. URL: <https://hal.archives-ouvertes.fr/hal-00302616>.

Cofiño, A.S., Bedia, J., Iturbide, M., Vega, M., Herrera, S., Fernández, J., Frías, M.D., Manzanas, R., 2018. The ECOMS User Data Gateway: towards seasonal forecast data provision and research reproducibility in the era of Climate Services. Clim. Serv. 9, 33–43.

Cohen, J., Jones, J., 2011. A new index for more accurate winter predictions. Geophys. Res. Lett. 38. http://dx.doi.org/10.1029/2011GL049626. n/a–n/a.

Collins, W.D., Rasch, P.J., Boville, B.A., Hack, J.J., McCaa, J.R., Williamson, D.L., Kiehl, J. T., Briegleb, B., Bitz, C., Lin, S.J., 2004. Description of the NCAR community atmosphere model (CAM 3.0). NCAR Techical Note NCAR/TN-464+ STR.

Cubasch, U., von Storch, H., Waszkewitz, J., Zorita, E., 1996. Estimates of climate change in Southern Europe derived from dynamical climate model output. Clim. Res. 7, 129–149. http://dx.doi.org/10.3354/cr007129. URL: <http://www.int-res.com/abstracts/cr/v07/n2/p129-149/>.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Holm, E.V., Isaksen, L., Kallberg, P., Koehler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.J., Park, B.K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J.N., Vitart, F., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137, 553–597. http://dx.doi.org/10.1002/qj.828.

Dickinson, R.E., Henderson-Sellers, A., Kennedy, P.J., 1993. Biosphere-Atmosphere Transfer Scheme (BATS) Version 1e as Coupled to the NCAR Community Climate Model. Technical Report NCAR/TN-387+STR. NCAR Technical Note. URL: <https://opensky.ucar.edu/islandora/object/technotes%3A154>.

Díez, E., Orfila, B., Frías, M.D., Fernández, J., Cofiño, A.S., Gutiérez, J.M., 2011. Downscaling ECMWF seasonal precipitation forecasts in Europe using the RCA model. Tellus Ser. A 63, 757–762. http://dx.doi.org/10.1111/j.1600-0870.2011.00523.x.

Díez, E., Primo, C., García-Moya, J.A., Gutiérrez, J.M., Orfila, B., 2005. Statistical and dynamical downscaling of precipitation over Spain from DEMETER seasonal forecasts. Tellus A 57, 409–423. http://dx.doi.org/10.1111/j.1600-0870.2005.00130.x.
Diro, G.T., 2016. Skill and economic benefits of dynamical downscaling of ECMWF ENSEMBLE seasonal forecast over southern Africa with RegCM4. Int. J. Climatol. 36, 675–688. http://dx.doi.org/10.1002/joc.4375.
Diro, G.T., Tompkins, A.M., Bi, X., 2012. Dynamical downscaling of ECMWF Ensemble seasonal forecasts over East Africa with RegCM3. J. Geophys. Res. Atmos. 117. http://dx.doi.org/10.1029/2011JD016997.
Doblas-Reyes, F.J., García-Serrano, J., Lienert, F., Biescas, A.P., Rodrigues, L.R.L., 2013. Seasonal climate predictability and forecasting: status and prospects. Wiley Interdiscip. Rev. 4, 245–268. http://dx.doi.org/10.1002/wcc.217.
ECWMF, 2007. IFS Documentation Cy31r1. Part IV: Physical Process, 155 pp. Technical Report. Shinfield Park, Reading, Berkshire RG2 9A, United Kingdom. URL: <https://www.ecmwf.int/sites/default/files/elibrary/2007/9221-part-iv-physical-processes.pdf>.
ECWMF, 2009. IFS Documentation Cy33r1. Part IV: Physical Process, 162 pp. Technical Report. Shinfield Park, Reading, Berkshire RG2 9A, United Kingdom. URL: <https://www.ecmwf.int/sites/default/files/elibrary/2009/9227-part-iv-physical-processes.pdf>.
Eden, J., Prodhomme, C., Dell'Aquila, A., 2014. Representation of the seasonal evolution of selected extreme events in Europe and implications for better representing key mechanisms and processes in s2d forecast systems. Technical Report D22.2. SPECS Deliverable. URL: <http://www.specs-fp7.eu/sites/default/files/u1/SPECSD22.2.pdf>.
Fouquart, Y., Bonnel, B., 1980. Computations of solar heating of the earth's atmosphere: a new parameterization. Beitrage zur Phys. Atmos. 53, 35–62.
Fowler, H.J., Blenkinsop, S., Tebaldi, C., 2007. Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. Int. J. Climatol. 27, 1547–1578. http://dx.doi.org/10.1002/joc.1556.
Frías, M.D., Herrera, S., Cofiño, A.S., Gutiérrez, J.M., 2010. Assessing the skill of precipitation and temperature seasonal forecasts in Spain: windows of opportunity related to ENSO events. J. Clim. 23, 209–220. http://dx.doi.org/10.1175/2009JCLI2824.1.
García-Díez, M., Fernández, J., Fita, L., Yagüe, C., 2013. Seasonal dependence of WRF model biases and sensitivity to PBL schemes over Europe. Q. J. R. Meteorol. Soc. 139, 501–514. http://dx.doi.org/10.1002/qj.1976.
García-Díez, M., Fernández, J., Vautard, R., 2015. An RCM multi-physics ensemble over Europe: multi-variable evaluation to avoid error compensation. Clim. Dyn. 1–16. http://dx.doi.org/10.1007/s00382-015-2529-x.
Giorgi, F., Coppola, E., Solmon, F., Mariotti, L., Sylla, M., Bi, X., Elguindi, N., Diro, G.T., Nair, V., Giuliani, G., Turuncoglu, U.U., Cozzini, S., Güttler, I., O'Brien, T.A., Tawfik, A.B., Shalaby, A., Zakey, A.S., Steiner, A.L., Stordal, F., Sloan, L.C., Brankovic, C., 2012. RegCM4: model description and preliminary tests over multiple CORDEX domains. Clim. Res. 52, 7–29. http://dx.doi.org/10.3354/cr01018. URL: <http://www.int-res.com/abstracts/cr/v52/p7-29/>.
Grell, G.A., Dudhia, J., Stauffer, D., 1994. A description of the fifth-generation Penn State/NCAR Mesoscale Model (MM5). Technical Report NCAR/TN-398+STR. NCAR Technical Note. URL: <http://opensky.ucar.edu/islandora/object/technotes:170>. doi:10.5065/D60Z716B.
Gutiérrez, J.M., Bedia, J., Benestad, R.C.P., 2013a. Review of the different statistical downscaling methods for s2d prediction. Technical Report. SPECS Deliverable D52, 1. URL: http://www.specs-fp7.eu/sites/default/files/u1/SPECS_D52.1.pdf.
Gutiérrez, J.M., San-Martín, D., Brands, S., Manzanas, R., Herrera, S., 2013b. Reassessing statistical downscaling techniques for their robust application under climate change conditions. J. Clim. 26, 171–188. http://dx.doi.org/10.1175/JCLI-D-11-00687.1.
Haylock, M.R., Cawley, G.C., Harpham, C., Wilby, R.L., Goodess, C.M., 2006. Downscaling heavy precipitation over the United Kingdom: a comparison of dynamical and statistical methods and their future scenarios. Int. J. Climatol. 26, 1397–1415. http://dx.doi.org/10.1002/joc.1318.
Haylock, M.R., Hofstra, N., Klein Tank, A.M.G., Klok, E.J., Jones, P.D., New, M., 2008. A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. J. Geophys. Res. 113. http://dx.doi.org/10.1029/2008JD010201.d20119. n/a–n/a.
Hazeleger, W., Severijns, C., Semmler, T., Stefanescu, S., Yang, S., Wang, X., Wyser, K., Dutra, E., Baldasano, J.M., Bintanja, R., Bougeault, P., Caballero, R., Ekman, A.M.L., Christensen, J.H., van den Hurk, B., Jimenez, P., Jones, C., Kållberg, P., Koenigk, T., McGrath, R., Miranda, P., Noije, T.V., Palmer, T., Parodi, J.A., Schmith, T., Selten, F., Storelvmo, T., Sterl, A., Tapamo, H., Vancoppenolle, M., Viterbo, P., Willén, U., 2010. EC-earth: a seamless earth-system prediction approach in action. Bull. Am. Meteorol. Soc. 91, 1357–1363. http://dx.doi.org/10.1175/2010BAMS2877.1.
Helsel, D.R., Hirsch, R.M., 2002. Statistical Methods in Water Resources. U.S Geological Survey.
Hertig, E., Jacobeit, J., 2008. Assessments of Mediterranean precipitation changes for the 21st century using statistical downscaling techniques. Int. J. Climatol. 28, 1025–1045. http://dx.doi.org/10.1002/joc.1597.
Hirschi, M., Seneviratne, S.I., Alexandrov, V., Boberg, F., Boroneant, C., Christensen, O.B., Formayer, H., Orlowsky, B., Stepanek, P., 2011. Observational evidence for soil-moisture impact on hot extremes in southeastern Europe. Nat. Geosci. 4, 17–21. http://dx.doi.org/10.1038/ngeo1032.
Hong, S., Noh, Y., Dudhia, J., 2006. A new vertical diffusion package with an explicit treatment of entrainment processes. Mon. Weather Rev. 134, 2318–2341. http://dx.doi.org/10.1175/MWR3199.1.
Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O.B., Bouwer, L., Braun, A., Colette, A., Déqué, M., Georgievski, G., Georgopoulou, E., Gobiet, A., Menut, L., Nikulin, G., Haensler, A., Hempelmann, N., Jones, C., Keuler, K., Kovats, S., Kröner, N., Kotlarski, S., Kriegsmann, A., Martin, E., van Meijgaard, E., Moseley, C., Pfeifer, S., Preuschmann, S., Radermacher, C., Radtke, K., Rechid, D., Rounsevell, M., Samuelsson, P., Somot, S., Soussana, J.F., Teichmann, C., Valentini, R., Vautard, R., Weber, B., Yiou, P., 2014. Euro-cordex: new high-resolution climate change projections for european impact research. Reg. Environ. Change 14, 563–578. http://dx.doi.org/10.1007/s10113-013-0499-2.
Kain, J.S., 2004. The Kain-Fritsch convective parameterization: an update. J. Appl. Meteorol. 43, 170–181. http://dx.doi.org/10.1175/1520-0450(2004) 0430170:TKCPAU2.0.CO;2.
Kang, I.S., Lee, J.Y., Park, C.K., 2004. Potential predictability of summer mean precipitation in a dynamical seasonal prediction system with systematic error correction. J. Clim. 17, 834–844. http://dx.doi.org/10.1175/1520-0442(2004) 0170834:PPOSMP2.0.CO;2.
Katragkou, E., Garća-Déz, M., Vautard, R., Sobolowski, S., Zanis, P., Alexandri, G., Cardoso, R.M., Colette, A., Fernández, J., Gobiet, A., Goergen, K., Karacostas, T., Knist, S., Mayer, S., Soares, P.M.M., Pytharoulis, I., Tegoulias, I., Tsikerdekis, A., Jacob, D., 2015. Regional climate hindcast simulations within EURO-CORDEX: evaluation of a WRF multi-physics ensemble. Geosci. Model Develop. 8, 603–618. http://dx.doi.org/10.5194/gmd-8-603-2015. URL: <http://www.geosci-model-dev.net/8/603/2015/>.
Kiehl, J.T., Hack, J.J., Bonan, G.B., Boville, B.A., Briegleb, B.P., Williamson, D.L., Rasch, P.J., 1996. Description of the NCAR Community Climate Model (CCM3). Technical Report NCAR/TN-420+STR. NCAR Technical Note. URL: <https://opensky.ucar.edu/islandora/object/technotes%3A187>.
Koster, R.D., Mahanama, S.P.P., Yamada, T.J., Balsamo, G., Berg, A.A., Boisserie, M., Dirmeyer, P.A., Doblas-Reyes, F.J., Drewitt, G., Gordon, C.T., Guo, Z., Jeong, J.H., Lawrence, D.M., Lee, W.S., Li, Z., Luo, L., Malyshev, S., Merryfield, W.J., Seneviratne, S.I., Stanelle, T., vanden Hurk, B.J.J.M., Vitart, F., Wood, E.F., 2010. Contribution of land surface initialization to subseasonal forecast skill: First results from a multi-model experiment. Geophys. Res. Lett., 37 http://dx.doi.org/10.1029/2009GL041677. n/a–n/a.
Kotlarski, S., Keuler, K., Christensen, O.B., Colette, A., Déqué, M., Gobiet, A., Goergen, K., Jacob, D., Lüthi, D., van Meijgaard, E., Nikulin, G., Schär, C., Teichmann, C., Vautard, R., Warrach-Sagi, K., Wulfmeyer, V., 2014. Regional climate modeling on european scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble. Geosci. Model Develop. 7, 1297–1333. http://dx.doi.org/10.5194/gmd-7-1297-2014.
Lachenbruch, P.A., Mickey, M.R., 1968. Estimation of error rates in discriminant analysis. Technometrics 10, 1–11. http://dx.doi.org/10.2307/1266219. URL: <http://www.jstor.org/stable/1266219>.
Laprise, R., 2008. Regional climate modelling. J. Comput. Phys. 227, 3641–3666. http://dx.doi.org/10.1016/j.jcp.2006.10.024. URL: <http://www.sciencedirect.com/science/article/pii/S0021999106005407>.
Lenderink, G., Holtslag, A.A.M., 2004. An updated length-scale formulation for turbulent mixing in clear and cloudy boundary layers. Q. J. R. Meteorol. Soc. 130, 3405–3427. http://dx.doi.org/10.1256/qj.03.117.
Lorenz, E.N., 1969. Atmospheric predictability as revealed by naturally occurring analogues. J. Atmos. Sci. 26, 636–646. URL: <http://dx.doi.org/10.1175/1520-0469(1969)26<636:APARBN>2.0.CO;2>.
Manzanas, R., Frías, M.D., Cofiño, A.S., Gutiérrez, J.M., 2014. Validation of 40 year multimodel seasonal precipitation forecasts: the role of ENSO on the global skill. J. Geophys. Res. 119, 1708–1719. http://dx.doi.org/10.1002/2013JD020680.
Manzanas, R., Lucero, A., Weisheimer, A., Gutiérrez, J.M., 2017. Can bias correction and statistical downscaling methods improve the skill of seasonal precipitation forecasts? Clim. Dyn. 1–16. http://dx.doi.org/10.1007/s00382-017-3668-z.
Maraun, D., 2016. Bias correcting climate change simulations – a critical review. Curr. Clim. Change Rep. 2, 211–220. http://dx.doi.org/10.1007/s40641-016-0050-x.
Maraun, D., Wetterhall, F., Ireson, A.M., Chandler, R.E., Kendon, E.J., Widmann, M., Brienen, S., Rust, H.W., Sauter, T., Themessl, M., Venema, V.K.C., Chun, K.P., Goodess, C.M., Jones, R.G., Onof, C., Vrac, M., Thiele-Eich, I., 2010. Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user. Rev. Geophys. 48, n/a–n/a. http://dx.doi.org/10.1029/2009RG000314.
Maraun, D., Widmann, M., Gutiérrez, J.M., Kotlarski, S., Chandler, R.E., Hertig, E., Wibig, J., Huth, R., Wilcke, R.A., 2015. VALUE: A framework to validate downscaling approaches for climate change studies. Earth's Future 3, 1–14. http://dx.doi.org/10.1002/2014EF000259.
Marzban, C., Sandgathe, S., Kalnay, E., 2006. MOS, perfect-prog and reanalysis. Mon. Weather Rev. 134, 657–663. http://dx.doi.org/10.1175/MWR3088.1.
van Meijgaard, E., van Ulft, L.H., Lenderink, G., deRoode, S.R., Wipfler, E.L., Boers, R., van Timmermans, R.M.A., 2012. Refinement and application of a regional atmospheric model for climate scenario calculations of Western Europe, 054/12. Technical Report. KNMI, PO Box 201, 3731 AE De Bilt, The Netherlands.
Mlawer, E.J., Taubman, S.J., Brown, P.D., Iacono, M.J., Clough, S.J., 1997. Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. J. Geophys. Res. 102, 16663–16682. http://dx.doi.org/10.1029/97JD00237.
Mo, R., Straus, D.M., 2002. Statistical-dynamical seasonal prediction based on Principal Component regression of GCM ensemble integrations. Mon. Weather Rev. 130, 2167–2187. http://dx.doi.org/10.1175/1520-0493(2002) 1302167:SDSPBO2.0.CO;2.

Murphy, J., 2000. Predictions of climate change over Europe using statistical and dynamical downscaling techniques. Int. J. Climatol. 20, 489–501. doi: 10.1002/(SICI)1097-0088(200004)20:5489::AID-JOC4843.0.CO, pp. 2-6.

Neggers, R.A.J., Koehler, M., Beljaars, A.A.M., 2009. A dual mass flux framework for boundary layer convection. Part I: Transport. J. Atmos. Sci. 66, 1465–1487. http://dx.doi.org/10.1175/2008JAS2635.1.

Nikulin, G., Asharafb, S., Magariño, M., Calmanti, S., Cardoso, R.M., Bhend, J., Fernández, J., Frías, M.D., Fröhlichb, K., Frühb, B., Herrera García, S., Manzanas, R., Gutiérrez, J.M., Hanssona, U., Kolaxa, M., Liniger, M., Soares, P.M., Spirig, C., Tome, R., Wysera, K., 2018. Dynamical and statistical downscaling of a global seasonal hindcast in Eastern Africa. Clim. Serv. 9, 72–85.

Nordeng, T.E., 1994. Extended versions of the convection parametrization scheme at ECMWF and their impact upon the mean climate and transient activity of the model in the tropics. Technical Report. Shinfield Park, Reading, Berkshire RG2 9A, United Kingdom.

Pal, J.S., Small, E.E., Eltahir, E.A.B., 2000. Simulation of regional-scale water and energy budgets: representation of subgrid cloud and precipitation processes within RegCM. J. Geophys. Res. 105, 29579–29594. http://dx.doi.org/10.1029/2000JD900415.

Patarčić, M., Branković, C., 2012. Skill of 2-m temperature seasonal forecasts over Europe in ECMWF and RegCM Models. Mon. Weather Rev. 140, 1326–1346. http://dx.doi.org/10.1175/MWR-D-11-00104.1.

Pavan, V., Marchesi, S., Morgillo, A., Cacciamani, C., Doblas-Reyes, F.J., 2005. Downscaling of DEMETER winter seasonal hindcasts over Northern Italy. Tellus A 57, 424–434. http://dx.doi.org/10.1111/j.1600-0870.2005.00111.x.

Preisendorfer, R., 1988. Principal component analysis in meteorology and oceanography. 1st ed., Elsevier.

Primo, C., Ferro, C.A., Jolliffe, I.T., Stephenson, D.B., 2009. Calibration of probabilistic forecasts of binary events. Mon. Weather Rev. 137, 1142–1149.

Robertson, A.W., Qian, J.H., Tippett, M.K., Moron, V., Lucero, A., 2012. Downscaling of seasonal rainfall over the philippines: dynamical versus statistical approaches. Mon. Weather Rev. 140, 1204–1218. http://dx.doi.org/10.1175/MWR-D-11-00177.1. URL: <http://journals.ametsoc.org/DOI/abs/10.1175/MWR-D-11-00177.1>.

Rust, H.W., Richling, A., Bissolli, P., Ulbrich, U., 2015. Linking teleconnection patterns to European temperature: a multiple linear regression model. Meteorol. Z. 24, 411–423. http://dx.doi.org/10.1127/metz/2015/0642.

Ruti, P.M., Somot, S., Giorgi, F., Dubois, C., Flaounas, E., Obermann, A., Dell'Aquila, A., Pisacane, G., Harzallah, A., Lombardi, E., Ahrens, B., Akhtar, N., Alias, A., Arsouze, T., Aznar, R., Bastin, S., Bartholy, J., Béranger, K., Beuvier, J., Bouffies-Cloché, S., Brauch, J., Cabos, W., Calmanti, S., Calvet, J.C., Carillo, A., Conte, D., Coppola, E., Djurdjevic, V., Drobinski, P., Elizalde-Arellano, A., Gaertner, M., Galàn, P., Gallardo, C., Gualdi, S., Goncalves, M., Jorba, O., Jordà, G., L'Heveder, B., Lebeaupin-Brossier, C., Li, L., Liguori, G., Lionello, P., Maciàs, D., Nabat, P., Önol, B., Raikovic, B., Ramage, K., Sevault, F., Sannino, G., Struglia, M.V., Sanna, A., Torma, C., Vervatis, V., 2016. Med-CORDEX initiative for mediterranean climate studies. Bull. Am. Meteorol. Soc. 97, 1187–1208. http://dx.doi.org/10.1175/BAMS-D-14-00176.1.

Sauter, T., Venema, V., 2011. Natural three-dimensional predictor domains for statistical precipitation downscaling. J. Clim. 24, 6132–6145. http://dx.doi.org/10.1175/2011JCLI4155.1.

Scaife, A.A., Arribas, A., Blockley, E., Brookshaw, A., Clark, R.T., Dunstone, N., Eade, R., Fereday, D., Folland, C.K., Gordon, M., Hermanson, L., Knight, J.R., Lea, D.J., MacLachlan, C., Maidens, A., Martin, M., Peterson, A.K., Smith, D., Vellinga, M., Wallace, E., Waters, J., Williams, A., 2014. Skillful long-range prediction of European and North American winters. Geophys. Res. Lett. 41, 2514–2519. http://dx.doi.org/10.1002/2014GL059637.

Schoof, J.T., Shin, D.W., Cocke, S., LaRow, T.E., Lim, Y.K., O'Brien, J.J., 2009. Dynamically and statistically downscaled seasonal temperature and precipitation hindcast ensembles for the southeastern USA. Int. J. Climatol. 29, 243–257. http://dx.doi.org/10.1002/joc.1717.

Sheridan, P., Smith, S., Brown, A., Vosper, S., 2010. A simple height-based correction for temperature downscaling in complex terrain. Meteorol. Appl. 17, 329–339. http://dx.doi.org/10.1002/met.177.

Shukla, S., Lettenmaier, D.P., 2013. Multi-RCM ensemble downscaling of NCEP CFS winter season forecasts: implications for seasonal hydrologic forecast skill. J. Geophys. Res. Atmos. 118, 10770–10790. http://dx.doi.org/10.1002/jgrd.50628.

Siebesma, A.P., Soares, P.M.M., Teixeira, J., 2007. A combined eddy-diffusivity mass-flux approach for the convective boundary layer. J. Atmos. Sci. 64, 1230–1248. http://dx.doi.org/10.1175/JAS3888.1.

Skamarock, W., Klemp, J., Dudhia, J., Gill, D., Barker, D., Duda, M., Wang, W., Powers, J., 2008. A Description of the Advanced Research WRF Version 3. Technical Report. NCAR.

Sokol, Z., 2003. MOS-based precipitation forecasts for river basins. Weather Forecast. 18, 769–781. http://dx.doi.org/10.1175/1520-0434(2003) 0180769:MPFFRB2.0.CO;2.

Spak, S., Holloway, T., Lynn, B., Goldberg, R., 2007. A comparison of statistical and dynamical downscaling for surface temperature in North America. J. Geophys. Res. 112, n/a–n/a. http://dx.doi.org/10.1029/2005JD006712.d08101.

von Storch, H., Zorita, E., Cubasch, U., 1993. Downscaling of global climate change estimates to regional scales: an application to iberian rainfall in wintertime. J. Clim. 6, 1161–1171. http://dx.doi.org/10.1175/1520-0442(1993) 0061161:DOGCCE2.0.CO;2.

Tiedtke, M., 1989. A comprehensive mass flux scheme for cumulus parameterization in large-scale models. Mon. Weather Rev. 117, 1779–1799. http://dx.doi.org/10.1175/1520-0493(1989) 1171779:ACMFSF2.0.CO;2.

Tiedtke, M., 1993. Representation of clouds in large-scale models. Mon. Weather Rev. 121, 3040–3061. http://dx.doi.org/10.1175/1520-0493(1993) 1213040:ROCILS2.0.CO;2.

Tompkins, A.M., Gierens, K., Rädel, G., 2007. Ice supersaturation in the ECMWF Integrated Forecast System. Q. J. R. Meteorol. Soc. 133, 53–63. http://dx.doi.org/10.1002/qj.14.

Undén, P., Järvinen, H., Rodríguez, E., Cats, G., 2003. Final Report of the HIRLAM-5 project. Technical Report. SMHI. SMHI, S-601 76 Norrköping, Sweden. URL: <http://www.hirlam.org/index.php/hirlam-documentation/catview/77-hirlam-official-publications/289-final-reports>.

Vannitsem, S., Nicolis, C., 2008. Dynamical properties of model output statistics forecasts. Mon. Weather Rev. 136, 405–419. http://dx.doi.org/10.1175/2007MWR2104.1.

Weisheimer, A., Palmer, T.N., 2014. On the reliability of seasonal climate forecasts. J. R. Soc. Interface 11. http://dx.doi.org/10.1098/rsif.2013.1162.

Wilby, R., Charles, S., Zorita, E., Timbal, B., Whetton, P., Mearns, L., 2004. Guidelines for use of climate scenarios developed from statistical downscaling methods. Technical Report.

Wilby, R.L., Wigley, T.M.L., 1997. Downscaling general circulation model output: a review of methods and limitations. Prog. Phys. Geogr. 21, 530–548. http://dx.doi.org/10.1177/030913339702100403. URL: <http://ppg.sagepub.com/content/21/4/530.abstract>.

Winkler, J.A., Guentchev, G.S., Perdinan, Tan, P.N., Zhong, S., Liszewska, M., Abraham, Z., Niedźwiedź, T., Ustrnul, Z., 2011. Climate scenario development and applications for local/regional climate change impact assessments: an overview for the non-climate scientist. Geogr. Compass 5, 275–300. http://dx.doi.org/10.1111/j.1749-8198.2011.00425.x.

Yoon, J.H., RubyLeung, L., Correia, J., 2012. Comparison of dynamically and statistically downscaled seasonal climate forecasts for the cold season over the United States. J. Geophys. Res., 117 http://dx.doi.org/10.1029/2012JD017650. URL: <http://onlinelibrary.wiley.com/DOI/10.1029/2012JD017650/abstract>.

Zorita, E., Hughes, J.P., Lettemaier, D.P., von Storch, H., 1995. Stochastic characterization of regional circulation patterns for climate model diagnosis and estimation of local precipitation. J. Clim. 8, 1023–1042. http://dx.doi.org/10.1175/1520-0442(1995) 0081023:SCORCP2.0.CO;2.

Zorita, E., von Storch, H., 1999. The analog method as a simple statistical downscaling technique: Comparison with more complicated methods. J. Clim. 12, 2474–2489. http://dx.doi.org/10.1175/1520-0442(1999) 0122474:TAMAAS2.0.CO;2.